Characterization of four wild edible Carduus species from the Mediterranean region via phytochemical and biomolecular analyses

Characterization of four wild edible *Carduus* species from the Mediterranean region *via* phytochemical and biomolecular analyses

Arianna Marengo[a,b,c], Andrea Maxia[a], Cinzia Sanna[a], Cinzia M. Bertea[b*], Carlo Bicchi[c], Mauro Ballero[a], Cecilia Cagliero[c], Patrizia Rubiolo[c*]

[a]Dipartimento di Scienze della Vita e dell'Ambiente, sezione di Botanica, Università di Cagliari, Viale Sant'Ignazio da Laconi 13, 09123 Cagliari, Italy

[b]Dipartimento di Scienze della Vita e Biologia dei Sistemi, Unità di Fisiologia Vegetale, Università di Torino, via Quarello 15/A, 10135 Torino, Italy

[c]Dipartimento di Scienza e Tecnologia del Farmaco, Università di Torino, Via P. Giuria 9, I-10125 Torino, Italy

[*]Corresponding authors

Prof. Dr. Patrizia Rubiolo

Dipartimento di Scienza e Tecnologia del Farmaco, Università di Torino, Via P. Giuria 9, I-10125 Torino, Italy, e-mail: patrizia.rubiolo@unito.it, tel:+39 011 670 7173 fax: +39 011 236 7661

Prof. Dr. Cinzia M Bertea

Unità di Fisiologia Vegetale, Dipartimento di Scienze della Vita e Biologia dei Sistemi, Università di Torino, via Quarello 15/A, 10135 Torino, Italy,

e-mail: cinzia.bertea@unito.it, tel: +39 011 670 6360 fax: +39 011 236 6360

Abstract

Carduus species (Compositae) are widely distributed in the Mediterranean area, and traditionally used for both food and medicinal purposes.

The hydroalcoholic extracts of four wild edible *Carduus* species collected in Sardinia (*Carduus argyroa* Biv., *Carduus nutans* subsp*. macrocephalus* (Desf.) Nyman, *Carduus pycnocephalus* L., *Carduus cephalanthus* Viv.) were analyzed and characterized by HPLC-PDA-MS/MS and PCR-RFLP of the nrDNA internal transcribed spacer (ITS).

Flavonoids and caffeoylquinic acid derivatives were the predominant classes of secondary metabolites characterizing the extracts. The ITS region was sequenced in parallel, and a PCR-RFLP method was applied with three selective restriction enzymes. Statistical analyses, on both chemical and biomolecular results, revealed that individuals clustered according to their taxonomic classification.

The combination of the two techniques discriminates the four species within the genus, giving further information on these little-investigated plants, traditionally used in the Mediterranean area and in Sardinia.

Keywords: *Carduus* spp., flavonoids, caffeoylquinic acids, HPLC-PDA-MS/MS, PCR-RFLP, ITS sequence

1. Introduction

Wild edible species are traditionally consumed mainly for their taste, as well as for their healthy and nutritional properties. Many popular dishes prepared with wild plants are still consumed, nowadays increasingly so with the return to traditions, with the primary aim of finding healthy alternatives to commercial foods (Guarrera & Savo, 2016). This is part of a trend aiming at re-discovering local products, often offered as culinary specialities in fairs and markets. Further, several edible plants are traditionally used as depuratives or to treat trivial illnesses. These species may potentially play an important role as functional foods, thanks to the great variety of physiologically-active components providing health benefits (Guarrera & Savo, 2016; Lentini & Venza, 2007; Pardo-de-Santayana et al., 2007; Ranfa, Maurizi, Romano, & Bodesmo, 2014). Species from the genus *Carduus* (Compositae family), known in English as thistles, are traditionally consumed for their taste and biological effects. They are annual or perennial plants, 0.5 -2 m high, with lance-shaped, spiny-toothed leaves, spiny-winged stems and white-to-purple flowers. The genus includes approximately 100 species worldwide, which are widely distributed over the Mediterranean area (Al-Shammari, Hassan, & Al-Youssef, 2015; Dimitrova-dyulgerova, Zhelev, & Mihaylova, 2015; Lahaye et al., 2008; Thao et al., 2011; Tutin et al., 1968). They are consumed as raw or cooked, and are used as medicinal plants for the treatment of liver disorders or, more in general, for their diuretic and digestive properties (Al-Shammari et al., 2015; Atzei, 2003; Dimitrova-dyulgerova et al., 2015; Guarrera & Savo, 2016; Lentini & Venza, 2007; Rinchen & Pant, 2014; Signorini, Piredda, & Bruschi, 2009; Tardío, Pardo-de-santayana, & Morales, 2006). Several classes of secondary metabolites, chiefly flavonoids, phenolic acids, lignans, coumarins, alkaloids, sterols, and triterpenes, have been found in these species (Al-Shammari et al., 2015; Cardona, García, José R., & Pérez, 1992; Dimitrova-dyulgerova et al., 2015; Fernández, Garcia, Pedro, & Varea, 1991; Jordon-Thaden & Louda, 2003). The presence of these compounds may be associated

to the documented wide range of biological and nutraceutical properties that are associated to *Carduus* species: liver tonicity, anti-inflammatory, antioxidant, antispasmodic, anticancer, antiviral, and antibacterial activity (Al-Shammari et al., 2015; Jordon-Thaden & Louda, 2003; Koc et al., 2015; Slavov, Mihayloiva, & Dimitrova-dyulgerova, 2014).

This study aims to verify similarities and dissimilarities in the chemical and biomolecular profiles of four wild edible *Carduus* species (*Carduus argyroa* Biv., *Carduus nutans* subsp. *macrocephalus* (Desf.) Nyman, *Carduus pycnocephalus* L., *Carduus cephalanthus* Viv) growing in the Mediterranean area and in particular in Sardinia, where these species are traditionally consumed. Little has been published on these species, in particular on those growing in Sardinia. Some data on the characteristic compounds (mainly flavonoids) of *C. pycnocephalus* and *C. nutans* L. (Al-Shammari et al., 2015; Bain & Desrochers, 1988; Jordon-Thaden & Louda, 2003; Marrelli, Loizzo, Nicoletti, Menichini, & Conforti, 2013) and on the polyacetylenes of *C. argyroa* extracts are available (Harborne, Baxter, & Moss, 1999; Jordon-Thaden & Louda, 2003). To the best of the authors' knowledge, no information is available on *C. cephalanthus* and *C. nutans* subsp. *macrocephalus*. At the same time, because of their relevance as traditional foods and remedies, it is of interest to learn more about these species, to verify the presence of compounds with nutraceutical properties, and to identify them in their extracts. The identification and discrimination of these closely-related species were approached by combining high performance liquid chromatography with diode array and mass spectrometry detectors (HPLC-PDA-MS/MS) and Restriction Fragment Length Polymorphism (PCR-RFLP) analysis of the nuclear ribosomal DNA (nrDNA) internal transcribed spacer (ITS) sequences, together with unsupervised multivariate data analysis and cluster analysis (PCA, HCA, Neighbor Joining, UPGMA). ITS gene as a DNA barcode marker is a useful tool to authenticate raw herbal materials, and in particular when (dried or processed) closely related species show similar chemical compositions. DNA barcoding has

successfully been applied to authenticate plant and animal samples, also in terms of food safety and quality control. However, a limit of this method is that reference sequences of uncommon plants are usually lacking in databases (Ali et al., 2014; Galimberti et al., 2013; Ha et al., 2015; Hebert, Cywinska, Ball, & Jeremy, 2003).

2. Materials and methods

2.1. Plant material

Aerial parts of the four wild species belonging to the genus *Carduus* were collected from different sites in Sardinia, in May and June 2015 (Table S1). They were identified at the Department of Life and Environmental Sciences, University of Cagliari, Italy, where a voucher specimen for each species was deposited. In total 10 specimens of *C. argyroa*, 6 of *C. cephalanthus*, 13 of *C. nutans* subsp*. macrocephalus* and 10 of *C. pycnocephalus* were collected. All plants growing at each site were separated by 1–50 m from one another, and were collected randomly. The fresh materials were dried at 40°C to constant weight.

2.2. Chemicals

HPLC-grade acetonitrile (LC-MS grade), formic acid (>98% purity), chlorogenic acid, rutin, apigenin 7-*O*-glucoside, apigenin, diosmin and kaempferol were from Sigma Aldrich (Bellefonte, USA). De-ionized water (18.2 MΩ cm) was obtained from a Milli-Q purification system (Millipore, Bedford, MA, USA). Luteolin, quercetin 3-*O*-glucoside, kaempferol 3-*O*-rutinoside and kaempferol 3-*O*-glucoside were from Extrasynthese (Genay Cedex, France). Cryptochlorogenic acid, 1,5-dicaffeoylquinic acid, 3,5-dicaffeoylquinic acid, 4,5-dicaffeoylquinic acid, luteolin 7-*O*-glucoside, luteolin 7-*O*-glucuronide, apigenin 7-*O*-glucuronide, kaempferol 3-*O*-rhamnoside, diosmetin and tricin were from Phytolab (Vestenbergsgreuth, Germany).

2.3. Sample preparation and HPLC-PDA-MS/MS analysis

Five hundred mg of each dried and ground aerial part were submitted to ultrasonic extraction with 10 mL of methanol/water (70:30, v/v) three times for 10 min. The extracts were then combined and centrifuged at 4000 rpm for 10 min. The supernatant was brought to a volume of 30 ml and filtered with a 13 mm diameter, 0.22 µm pore diameter hydrophilic PTFE syringe filter, before the HPLC-PDA-MS/MS analysis.

2.4. HPLC-PDA-MS/MS analysis

Each extract (5 µl) was analyzed in duplicate with a Shimadzu Nexera X2 system equipped with a photodiode detector SPD-M20A in series to a triple quadrupole Shimadzu LCMS-8040 system provided with electrospray ionization (ESI) source (Shimadzu, Dusseldorf Germany). Samples were analyzed on an Ascentis Express C18 column (15 cm x 2.1 mm, 2.7µm, Supelco, Bellefonte, USA) using water/formic acid (999:1, v/v) and acetonitrile/formic acid (999:1, v/v) as mobile phases A and B, respectively. The flow rate was 0.4 mL/min and the column temperature was maintained at 30°C. The gradient program was as follows: 5% B for 3 min, 5-15% B in 17 min, 15-25% B in 10 min, 25-75% B in 12 min, 75-100% B in 10 min, 100% B for 1 min. Total pre-running and post-running time was 60 min. UV spectra were acquired over the 220-450 nm wavelength range and the resulting chromatograms were integrated at 330 nm. MS operative conditions were as follows: heat block temperature: 200°C; desolvation line (DL) temperature: 250°C; nebulizer gas flow rate: 3 L/min drying gas flow rate: 15 L/min. Mass spectra were acquired both in positive and in negative full-scan mode over the range 100-1000 m/z, event time 0.5 sec. Product Ion Scan mode (collision energy: - 35.0 V for ESI^+ and 35.0 V for ESI^-, event time: 0.2 sec) was applied to compounds for which a correspondence between the pseudomolecular ions $[M+H]^+$ in ESI^+ and $[M-H]^-$ in ESI^- had been confirmed. Multiple Reaction Monitoring acquisition (collision energy: - 35.0 V for ESI^+ and 35.0 V for ESI^-, dwell time: 20) was carried out on specific product ions derived

from precursor ions fragmentation (Table S2). Some of the main components were identified by comparing their retention times, UV and MS spectra to those of authentic standards (chlorogenic acid, cryptochlorogenic acid, dicaffeoylquinic acids, rutin, quercetin 3-*O*-glucoside, luteolin 7-*O*-glucoside, luteolin 7-*O*-glucuronide, kaempferol 3-*O*-rutinoside, kaempferol 3-*O*-glucoside, kaempferol 7-*O*-rhamnoside, apigenin 7-*O*-glucoside, apigenin 7-*O*-glucuronide, diosmin, apigenin, luteolin, kaempferol, tricin, diosmetin). The other components were tentatively identified on the basis of their UV spectra and mass spectral information, compared to those given in the literature. Data were processed using LabSolution software (Shimadzu, Dusseldorf Germany).

2.5. DNA extraction, PCR amplification and sequencing

Ten milligrams of the same material employed for chemical analyses were ground to a fine powder, with the addition of approximately 5 mg of polyvinylpolypyrrolidone (PVPP, Sigma Aldrich, Bellefonte, USA). Genomic DNA was extracted from the ground powder using the Eurogold Plant DNA Mini Kit (Euroclone, Pero, Italy) following the manufacturer's instructions. The quantitative and qualitative analyses of the isolated genomic DNA were assessed by spectrophotometry using the Nanophotometer (Implen GmbH, Munich, Germany) and by gel electrophoresis. Approximately 20 ng of genomic DNA were used as a template for PCR amplification with forward primer ITS1 (5'-TCCGTAGGTGAACCTGCGG-3') and reverse primer ITS4 (5'-TCCTCCGCTTATTGATATGC-3') (White, Bruns, Lee, & Taylor, 1990). The amplification was carried out in a 25 µl reaction mixture containing 2.5 µl of 10X PCR buffer (Thermo-Scientific, Waltham, MA USA), 0.2 mM deoxynucleotide triphosphates (dNTPs), 20 pmol of forward and reverse primers, and 0.5 U of *Taq* DNA polymerase (Thermo-Scientific, Waltham, MA USA). PCR reactions were carried out in a T-Gradient Thermalcycler (Biometra GmbH, Göttingen, Germany). Cycling conditions consisted of an initial 4 min at 94°C, followed by 30 s of denaturing at 94°C, 45 s of annealing at 53°C and 45 s of elongation at 72°C, repeated for 35 cycles and with 10 min of final

extension at 72°C. PCR products were separated by 1.5% (w/v) agarose gel electrophoresis and visualized by ethidium bromide staining under UV. PCR products were purified using the Agencourt® AMPure® kit following the manufacturer's protocol, and employed as a template for sequencing (IGA Technology Services, Udine, Italy). Both DNA strands were sequenced.

2.6. PCR-RFLP

The PCR products of the ITS gene of the four species were digested in separate reactions with 10 U of *Nde*I at 37°C for 2 h, or with 10U of *Stu*I or of *Apa*LI, at 37°C for 1 h (NEB, New England Biolabs, Ipswich, AM, USA). One microliter of each digestion reaction was analyzed by capillary gel electrophoresis (CGE) using the Agilent 2100 Bioanalyzer (Agilent Technologies) and the DNA 1000 LabChip Kit (Agilent Technologies) following the manufacturer's instructions.

The DNA 1000 LabChip Kit provides sizing and quantitation of dsDNA fragments ranging from 25 to 1000 bp.

2.7. Statistical elaboration

Data obtained from the chemical analyses were processed through Principal Component Analysis (PCA), to reduce the multivariate space in which objects were distributed, and through hierarchical cluster analysis with quadratic Euclidean distance and Ward linkage, using SPSS 15.0 (IBM Corporation) software. Before data treatment, all variables obtained from HPLC-PDA-MS/MS analysis were scaled to unit variance.

Gene sequences were aligned with CLC sequence viewer software using default parameters to check the integrity of each sample sequence. Consensus sequences, obtained by aligning the individual sequences of each species, were then aligned using MEGA7 software (ClustalW) by modifying the Gap Opening and Gap Extension Cost values to 15 and 1, respectively. From this last alignment, a cluster analysis was made. Neighbor Joining and UPGMA statistical methods were

selected and relationships were tested with 1000 Bootstrap replicates, considering gaps in the Partial Deletion option.

3. Results and discussion

3.1. Phytochemical analyses

The extracts of *Carduus argyroa*, *Carduus cephalanthus*, *Carduus nutans* subsp. *macrocephalus* and *Carduus pycnocephalus* were analyzed by HPLC-PDA-MS/MS, to obtain their chromatographic profiles, and to obtain UV spectra, and mass spectral information concerning their components. MS data were acquired for a significant number of plants per species to identify the secondary metabolites characterizing them; all plants were analyzed twice. The repeatability of extract composition was evaluated on at least three plants per species. Fig.1 shows a representative chromatogram for each species. Approximately sixty peaks were detected through the HPLC-PDA-MS/MS untargeted metabolite analysis. Accordingly to the UV and MS collected information, 31 informative peaks were identified or tentatively identified and selected as target compounds for the statistical analysis. Among them, 20 were identified by comparing their UV and MS spectra to those of reference standards. The other 11 components were tentatively identified from their UV, MS and MS/MS spectra by comparison to literature data (Table 1). In agreement with the existing literature, flavonoids were the most representative compounds, and in particular quercetin, luteolin, kaempferol, diosmetin, and apigenin *O*-glycosides, resulting in 18 flavones and 7 flavonols; in addition, caffeoylquinic acids were also found (Al-Shammari et al., 2015; Ha et al., 2015; Jordon-Thaden & Louda, 2003; Li et al., 2014; Thao et al., 2011). Tandem mass spectrometry fragmentation provides further structural information on the compounds for which authentic standards are not available. The following compounds are given as illustrative examples of how the fragmentation patterns were used for identification. Compounds **4**, **5**, **6**, **9**, **10** and **16** are

characterized by maximum UV absorptions at approximately 348, 253 and 265 nm, which are typical for flavones (Li et al., 2014; Pandino, Lombardo, Mauromicale, & Williamson, 2011). For each peak, the MS/MS fragmentation gave a potential aglycone of 286 g/mol, and the positive ion mode MS/MS fragmentation yielded $^{1,3}B^+$ (*m/z* 135) and $^{1,3}A^+$ (*m/z* 153) fragments due to the $X^{1,3}$ cleavage of the C-ring, $^{0,2}B^+$ (*m/z* 137) and $^{0,4}B^+$-H_2O (*m/z* 161) fragments due to the $X^{0,2}$ and $X^{0,4}$ cleavage of the C-ring, respectively. Since this fragmentation pattern is consistent with that of luteolin standard (compound **26**), compounds **4**, **5**, **6**, **16** were tentatively identified as luteolin O-glycosides. Compounds **9** and **10** were identified as luteolin glucuronide and luteolin glucoside, respectively, by comparison with the authentic standards. (Cuyckens & Claeys, 2004). Apigenin (**17**, **18**, **20**), kaempferol (**12**, **14**, **24**), diosmetin (**19**, **21**, **22**, **25**, **27**), quercetin (**3**, **7**, **8**), and their glycosides were identified or tentatively identified in the extracts with the same approach. The pseudomolecular $[M - H]^+$ diagnostic fragments of the aglycones were as follows: apigenin: MS^2 at *m/z* 119, 153, 163; kaempferol: MS^2 at *m/z* 121, 137, 153, 165; diosmetin: MS^2 at *m/z* 258, 286; quercetin: MS^2 at *m/z* 121, 127, 137, 153, 161, 165; tricin: MS^2 at *m/z* 153, 203, 315. Six caffeoylquinic acid derivatives were also identified (**1**, **2**, **11**, **13**, **15**, **23**), by comparison with reference standards. The only exception was compound **11**, for which the MS/MS fragments (*m/z* 163 and 191 in the positive and negative ionization mode, respectively) were consistent with those diagnostic for the dicaffeoylquinic acid derivatives (Li et al., 2014; Martini, Conte, & Tagliazucchi, 2017).

From these results, interspecific differences can be associated to the presence/absence of specific compounds (Table 1). Some compounds, such as chlorogenic acid (**1**), cryptochlorogenic acid (**2**), dicaffeoylquinic acids **(13** and **14**), kaempferol 3-*O*-glucoside (**15**), kaempferol 3-*O*-rhamnoside (**24**), luteolin (**26**), apigenin (**28**), kaempferol (**29**), diosmetin (**30**) and tricin (**31**), were present in all samples. Diosmetin derivatives **(19**, **21**, **22**) and 4,5 dicaffeoylquinic acid (**23**) were only detected in

C. cephalanthus and *C. pycnocephalus,* while luteolin *O*-arabinosyl-glucoside (**5**), apigenin *O*-rhamnosyl-glucoside (**17**), and apigenin 7-*O*-glucoside (**18**) were only present in *C. argyroa* and *C. nutans* subsp*. macrocephalus* samples. Moreover, some peaks, detectable only in one of the four species, can be taken as markers. In particular, luteolin *O*-arabinosyl-glucoside isomer (**6**) is present only in *C. argyroa*, dicaffeoylquinic acid (**11**) in *C. cephalanthus*, luteolin diglucoside (**4**), luteolin acetyl diglycoside (**16**) and diosmetin acetyl glycosides (**25**, **27**) in *C. nutans* subsp*. macrocephalus,* and a quercetin diglycoside (**3**) in *C. pycnocephalus*. In spite of the same retention time (19.898 min) and pseudomolecular ions (611 in ESI^+ and 609 in ESI^- , i.e. supposed molecular weight 610 g/mol), compounds **3** and **4** fragmented differently in MS/MS, giving potential aglycones at 302 g/mol and 286 g/mol, respectively.

A survey of the literature confirmed some of the results, viz. the presence of chlorogenic acid, rutin, kaempferol, apigenin, luteolin and the related *O*-glycosides in *C. pycnocephalus* and *C. nutans* (Al-Shammari et al., 2015; Bain & Desrochers, 1988; Jordon-Thaden & Louda, 2003; Marrelli et al., 2013). The other (putatively) identified compounds, namely dicaffeoylquinic acids, diosmetin, quercetin, and tricin, were found in other *Carduus* species (Dimitrova-dyulgerova et al., 2015; Jordon-Thaden & Louda, 2003; Li et al., 2014). No data are available on the flavonoid composition of *C. argyroa*, *C. cephalanthus* and *C. nutans* subsp. *macrocephalus*.

It is worth noting that the compounds detected in the species investigated are known to have several nutraceutical properties. For instance, the antioxidant, hepatoprotective, antibacterial, anticarcinogenic, and anti-inflammatory activities of most of the flavonoids and caffeoylquinic acid derivatives identified are documented (de Falco, Incerti, Amato, & Lanzotti, 2015; Nijveldt et al., 2001).

3.2. Statistical data treatment

Principal Component Analysis (PCA) and hierarchical cluster analysis (HCA), were applied to discriminate samples from the species investigated. PCA is an unsupervised multivariate data analysis method with which the multivariate space in which objects are distributed can be reduced, so as to visualize similarities and/or differences within multivariate data of secondary metabolite composition (Zheng, Jiang, Wu, Wang, & Huang, 2014). HCA is an unsupervised pattern recognition method to detect similarities (Ha et al., 2015). The dataset of the 39 samples investigated was thus submitted to PCA and HCA, in a targeted approach based on the 31 previously characterized compounds listed in Table 1. The best result was obtained by scaling all variables to unit variance, so that all components had a standard deviation of one (Berg, Hoefsloot, Westerhuis, Smilde, & Werf, 2006).

Both score and loading plots were built with the aim of discriminating the four species and verifying the presence of discriminating variables. As shown in Fig. 2A, the first component (PC1) that explains 42.97% of the variation, individually separates *C. argyroa* and *C. nutans* subsp*. macrocephalus* from *C. pycnocephalus* and *C. cephalanthus*. Conversely, PC2 (20.98% of the variation) discriminates between *C. argyroa* and C*. nutans* subsp*. macrocephalus.* The related loading plot shows the influence of the variables in the distribution of the samples in the loading plot (Fig. 2B). From the results obtained in this plot, it is interesting to note that, among the compounds that negatively explain PC1, **19**, **21**, **22**, **23**, are present only in *C. pycnocephalus* and *C. cephalanthus* samples, while compounds **5**, **17**, and **18** are positively correlated with PC1 and are characteristic of *C. argyroa* and *C. nutans* subsp *macrocephalus* species (Table 1). Conversely, among the PC2 positively correlated variables, compounds **4**, **25** and **27** are characteristic of *C. nutans* subsp*. macrocephalus* while **6**, which is negatively correlated with PC2, is only present in *C. argyroa* (Fig. 2B). The third component (PC3) explains 10.30% of the variation and separates *C. cephalanthus* from *C. pycnocephalus* (Fig.2C). Fig. 2D highlights the positive correlation of **3** and

the negative correlation of **11** with PC3; these compounds are characteristic of *C. pycnocephalus* and *C. cephalanthus* species, respectively (Table 1).

These results suggest that some of these compounds can be discriminating for species identification purposes, although all components must be considered together for better species discrimination.

A three-dimensional PCA was also carried out, since the combination of the first two components is not sufficient to separate all individuals belonging to each species. The first three components, which explain 74.25% of the variation, give four distinct clusters according to species (Fig. S1).

In the next step, a dendrogram generated by hierarchic cluster analysis was constructed. Fig. 2E shows that all samples belonging to the same species are clustered together. This dendrogram confirms the results obtained with the PCA, and provides more information on the relationships among the four species. In particular, the chemical composition of *C. argyroa* appears to be more closely correlated to that of *C. nutans* subsp. *macrocephalus*, while *C. pycnocephalus* and *C. cephalanthus* cluster together.

3.3 ITS sequence analysis

For the PCR amplification, primers flanking the ITS-1 and ITS-2 regions were employed (Fig. 3A). It has been shown that the ITS regions provide better discrimination at the species level, for species belonging to Compositae family (Choi & Thines, 2015). This sequence comprises the rDNAs transcription units, which are well conserved in the higher plants, and the ITS-1 and ITS-2 regions, which vary widely in closely related species (Wu et al., 2011). The nucleotide composition of the resulting sequences is shown in Fig. 3B. The sequences are approximately 785-811 bp long, and the alignment of the four species' sequences shows that 89% of the sites are conserved. Data suggest the presence of some differences in the nucleotide composition of the sequences belonging to each species. In particular, of the 10.3% that are variable sites, 0.6% provide little

information and 9.3% are singleton sites. In agreement with the literature, most of these variable sites are located in the ITS-1 and ITS-2 regions (Wu et al., 2011). Analysis of the individual ITS sequences indicates the lack of intraspecific nucleotide variation (data not shown). A consensus sequence was built for each species.

The Neighbor Joining (NJ) and the Unweighted Pair Group Method with Arithmetic Mean (UPGMA) trees were built from the ITS region sequence alignment (Fig. 4 A, B). Cluster analyses showed a cluster linking *C. argyroa* to *C. nutans* subsp*. macrocephalus.* The *C. cephalanthus* nucleotide composition appeared to vary more significantly than the sequences of the other species. These results are in agreement with the dendrogram based on chemical data (Fig. 2E).

Analysis of the sequences obtained through pairwise sequence alignment with the software BLASTN 2.4.0+ gave the percentage of identity between each pair of sequences (Table S3) (Zhang, Schwartz, Wagner, & Miller, 2000). From the resulting percentages of identity, it emerged that the *C. nutans* subsp*. macrocephalus* sequence is more similar to those of *C. argyroa* and *C. pycnocephalus* (95% and 94%, respectively). *C. argyroa* and *C. pycnocephalus* showed 94% nucleotide identity, while the *C. cephalanthus* sequence was more variable than the sequences of the other species (92-93% of identity vs. the other species). These data are in accordance with the results of the hierarchical cluster analysis (Fig. 4).

A further interesting aspect is that each sequence is species-specific, and can be used as a barcoding gene. ITS sequences of *C. pycnocephalus*, *C. nutans* and *C. nutans* subsp*. leiophyllus* (Petrovič) Stoj. & Stef. and other *Carduus* species are deposited in GenBank (Kelch & Baldwin, 2003; Robba, Carine, Russell, & Raimondo, 2005; Susanna, Hidalgo, Vilatersana, & Ciencia, 2006). The *Carduus* database ITS sequences, which include the ITS-1, 5.8 rRNA gene, and ITS-2 complete sequences, and 18S and 28S ribosomal RNA gene partial sequences, are generally shorter (from 686 to 737 bp) than those reported here. Blast analysis shows that the *C. pycnocephalus* sequence

obtained here is identical to those present in the database (EF123105.1, KT363916.1). Additionally, the *C. nutans* subsp*. macrocephalus* sequence is almost identical to those of *C. nutans* and *C. nutans* subsp*. leiophyllus* (98-99% identity) (AF443678.1, HQ540426.1, EF543521.1, KT249753.1, KC603920.1, KT363914.1, JX867642.1, KX167785.1). In particular, the *C. nutans* subsp*. leiophyllus* ITS sequence shows only one nucleotide variation versus the *C. nutans* ITS sequence. Conversely, in positions 70, 298 and 627, the *C. nutans* subsp*. macrocephalus* sequence has an adenine, while the *C. nutans* sequences already reported have a guanine, a cytosine, and a thymine, in the respective positions. No sequences belonging to the species *C. argyroa* and *C. cephalanthus* are present in GenBank, but BLAST alignment indicates the similarity of these sequences to those belonging to other *Carduus* species. This reveals that the ITS sequence has some conserved regions throughout the genus *Carduus,* even if variable nucleotide sites can discriminate between the individual species. The fact that the *C. nutans* subsp*. macrocephalus* sequence is almost identical to those of *C. nutans* and *C. nutans* subsp*. leiophyllus* may suggest that, in this case, this gene is not discriminatory at the subspecies level.

3.4. PCR-RFLP analysis

To better discriminate among the species, PCR-RFLP analyses were carried out. Three restriction enzymes were used to selectively cleave the resulting amplicons (Fig. 3C). The first four lanes of Fig. 5 show the undigested PCR products of the four *Carduus* species. From the RFLP analysis, it is possible to note that *Nde*I, ineffective on the other *Carduus* sequences, is able to discriminate *C. nutans* subsp. *macrocephalus* by giving two distinct fragments (512 and 295 bp, respectively). Conversely, *Stu*I selectively cleaves only *C. argyroa* and *C. pycnocephalus* sequences, giving two diagnostic fragments (623 and 188 bp for *C. argyroa*, and 620 and 178 bp for *C. pycnocephalus*). Since the *C. argyroa* and *C. pycnocephalus* sequences are cleaved by *Stu*I at the same sites, *Apa*LI was employed to discriminate between these two species. This enzyme is able to cleave only *C.*

pycnocephalus, releasing two diagnostic fragments of 508 and 290 bp, respectively. Lastly, the *C. cephalanthus* sequences show no cleavage sites for the three enzymes analyzed. These results show that it is possible to differentiate among the four species investigated with a combination of these three different restriction enzymes.

4. Conclusion

This study reports the first investigation of the phenolic acid and flavonoid profiles, and the ITS sequences and PCR-RFLP, of *C. argyroa, C. cephalanthus, C. pycnocephalus* and *C. nutans* subsp. *macrocephalus* from Sardinia. The differences in secondary metabolite profiles, defined by HPLC-PDA-MS/MS analysis, together with Principal Components and Hierarchical Clustering Analysis, afforded their unequivocal discrimination. Similar results were obtained with biomolecular analysis, through ITS sequence and PCR-RFLP analyses. ITS sequences of *C. pycnocephalus*, *C. nutans* and *C. nutans* subsp. *leiophyllus* and other *Carduus* species are available in the database (Kelch & Baldwin, 2003; Robba, Carine, Russell, & Raimondo, 2005; Susanna, Hidalgo, Vilatersana, & Ciencia, 2006) while the ITS genes of *C. argyroa*, *C. cephalanthus* and *C. nutans* subsp. *macrocephalus* were here sequenced for the first time and deposited in GenBank.

The complementary combination of the genetic and chemical approaches provides important information on little-investigated, but traditionally widely-used, plants and offers reliable discrimination of four morphologically-similar species belonging to the same genus. Moreover, chemical analysis detected interesting compounds with nutraceutical properties in the extracts investigated. This might support the traditional consumption and medicinal uses of these plants, and open new perspectives for further investigation of the compound(s) potentially responsible for the properties attributed to them. These findings may also promote the consumption of these herbs, including through the development of food supplements and functional foods. Moreover,

this approach can successfully be used for quality control of the species, e.g. in thistle-based commercial food products or traditional herbal remedy, for which their botanical and chemical composition must be quoted to assess origin, or because of regulatory requirements (e.g. food supplements).

Funding

This work was in part supported by Local Research Grants from the University of Turin, Italy (to CMB).

References

Al-Shammari, L. A., Hassan, W. H. B., & Al-Youssef, H. M. (2015). Phytochemical and biological studies of *Carduus pycnocephalus* L. *Journal of Saudi Chemical Society*, *19*, 410–416.

Ali, M. A., Gyulai, G., Hidvégi, N., Kerti, B., Al Hemaid, F. M. A., Pandey, A. K., & Lee, J. (2014). The changing epitome of species identification – DNA barcoding. *Saudi Journal of Biological Sciences*, *21*, 204–231.

Atzei, A. D. (2003). Le piante nella tradizione popolare della Sardegna. *Carlo Delfino Editore*.

Bain, J. F., & Desrochers, A. M. (1988). Flavonoids of *Carduus nutans* and *C. acanthoides*. *Biochemical Systematics and Ecology*, *16*(3), 265–268.

Berg, R. A. Van Den, Hoefsloot, H. C. J., Westerhuis, J. A., Smilde, A. K., & Werf, M. J. Van Der. (2006). Centering, scaling, and transformations: improving the biological information content of metabolomics data. *BMC Genomics*, *7*, 142–157.

Cardona, L., García, B., José R., P., & Pérez, J. (1992). 6-Prenyloxy-7-methoxycoumarin, a coumarin-hemiterpene ether from *Carduus tenuiflorus*. *Phytochemistry*, *31*, 3989–3991.

Choi, Y. J., & Thines, M. (2015). Host jumps and radiation, not co - Divergence drives diversification

of obligate pathogens. A case study in downy mildews and Asteraceae. *PLoS ONE*, *10*. https://doi.org/10.1371/journal.pone.0133655

Cuyckens, F., & Claeys, M. (2004). Mass spectrometry in the structural analysis of flavonoids. *Journal of Mass Spectrometry*, *39*, 1–15.

de Falco, B., Incerti, G., Amato, M., & Lanzotti, V. (2015). Artichoke: botanical, agronomical, phytochemical, and pharmacological overview. *Phytochemistry Reviews*, *14*, 993–1018.

Dimitrova-dyulgerova, I., Zhelev, I., & Mihaylova, D. (2015). Phenolic profile and in vitro antioxidant activity of endemic Bulgarian *Carduus* species. *Pharmacognosy Magazine*, *11*, 575–579.

Fernández, I., Garcia, B., Pedro, J. R., & Varea, A. (1991). Lignans and flavonoids from *Carduus assoi*. *Phytochemistry*, *30*, 1030–1032.

Galimberti, A., Mattia, F. De, Losa, A., Bruni, I., Federici, S., Casiraghi, M., … Labra, M. (2013). DNA barcoding as a new tool for food traceability. *Food Research International*, *50*, 55–63.

Guarrera, P. M., & Savo, V. (2016). Wild food plants used in traditional vegetable mixtures in Italy. *Journal of Ethnopharmacology*, *185*, 202–234.

Ha, I. J., Lee, M. Y., Kwon, Y.-K., Jung, Y., Kim, H. K., & Hwang, G.-S. (2015). Metabolite profiling to discriminate different species and genus from thistles in Korea using liquid chromatography with quadrupole time-of-flight mass spectrometry. *Journal of Separation Science*, *38*(3), 502–510.

Harborne, J. B., Baxter, H., & Moss, G. P. (1999). Phytochemical Dictionary, A Handbook of Bioactive Compounds from Plants, 2nd edn. *Taylor and Francis, London.*

Hebert, P. D. N., Cywinska, A., Ball, S. L., & Jeremy, R. (2003). Biological identifications through DNA barcodes. *Proceedings of the Royal Society of London*, *270*, 313–321.

Jordon-Thaden, I. E., & Louda, S. M. (2003). Chemistry of *Cirsium* and *Carduus*: A role in ecological

risk assessment for biological control of weeds? *Biochemical Systematics and Ecology*, *31*(12), 1353–1396.

Koc, S., Isgor, B. S., Isgor, Y. G., Shomali Moghaddam, N., Yildirim, O., Shomali, N., … Yildirim, O. (2015). The potential medicinal value of plants from Asteraceae family with antioxidant defense enzymes as biological targets. *Pharmaceutical Biology*, *53*(5), 746–51.

Lahaye, R., Bank, M. Van Der, Bogarin, D., Warner, J., Pupulin, F., Gigot, G., … Savolainen, V. (2008). DNA barcoding the floras of biodiversity hotspots. *PNAS*, *105*, 2923–2928.

Lentini, F., & Venza, F. (2007). Wild food plants of popular use in Sicily. *Journal of Ethnobiology and Ethnomedicine*, *3*. https://doi.org/10.1186/1746-4269-3-15

Li, R., Liu, S. K., Song, W., Wang, Y., Li, Y. J., Qiao, X., … Ye, M. (2014). Chemical analysis of the Tibetan herbal medicine *Carduus acanthoides* by UPLC/DAD/qTOF-MS and simultaneous determination of nine major compounds. *Analytical Methods*, *6*(18), 7181–7189.

Marrelli, M., Loizzo, M. R., Nicoletti, M., Menichini, F., & Conforti, F. (2013). Inhibition of Key Enzymes Linked to Obesity by Preparations From Mediterranean Dietary Plants: Effects on α-Amylase and Pancreatic Lipase Activities. *Plant Foods for Human Nutrition*, *68*(4), 340–346.

Martini, S., Conte, A., & Tagliazucchi, D. (2017). Phenolic compounds profile and antioxidant properties of six sweet cherry (*Prunus avium*) cultivars. *Food Research International*, *97*, 15–26.

Nijveldt, R. J., van Nood, E. van, van Hoorn, D. E., Boelens, P. G., van Norren, K., & van Leeuwen, P. A. (2001). Flavonoids: a review of probable mechanism of action and potential applications. *The American Journal of Clinical Nutrition*, *74*, 418–425.

Pandino, G., Lombardo, S., Mauromicale, G., & Williamson, G. (2011). Profile of polyphenols and phenolic acids in bracts and receptacles of globe artichoke (*Cynara cardunculus* var . *scolymus*) germplasm. *Journal of Food Composition and Analysis*, *24*, 148–153.

Pardo-de-Santayana, M., Tardío, J., Blanco, E., Carvalho, A., Lastra, J., San Miguel, E., & Morales, R. (2007). Traditional knowledge of wild edible plants used in the northwest of the Iberian Peninsula (Spain and Portugal): a comparative study. *Journal of Ethnobiology and Ethnomedicine*, *3*, doi:10.1186/1746-4269-3-27.

Ranfa, A., Maurizi, A., Romano, B., & Bodesmo, M. (2014). The importance of traditional uses and nutraceutical aspects of some edible wild plants in human nutrition: the case of Umbria (central Italy). *Plant Biosystems*, *148*(2), 297–306.

Rinchen, T., & Pant, S. (2014). Ethopharmacological uses of plants among inhabitants surronding Suru and Zanskar valleys of cold desert, Ladakh. *International Journal of Pharma and Bio Sciences*, *5*, 486–494.

Signorini, M. A., Piredda, M., & Bruschi, P. (2009). Plants and traditional knowledge: an ethnobotanical investigation on Monte Ortobene (Nuoro, Sardinia). *Journal of Ethnobiology and Ethnomedicine*, *5*(6), 1–14.

Slavov, I., Mihayloiva, D., & Dimitrova-dyulgerova, I. (2014). Phenolic acids, flavonoid profile and antioxidant activity of carduus thoermeri Weinm. Extract. *Oxidation Communications*, *37*, 247–253.

Tardío, J., Pardo-de-santayana, M., & Morales, R. (2006). Ethnobotanical review of wild edible plants in Spain. *Botanical Journal of the Linnean Society*, *152*, 27–71.

Thao, N. T. P., Cuong, T. D., Hung, T. M., Lee, J. H., Na, M., Son, J. K., … Min, B. S. (2011). Simultaneous determination of bioactive flavonoids in some selected korean thistles by high-performance liquid chromatography. *Archives of Pharmacal Research*, *34*, 455–461.

Tutin, T. G., Heywood, V. H., Burges, N. A., Valentine, D. H., Walters, S. M., & Webb, D. A. (1968). Flora europaea. *Cambridge University Press*.

White, T. J., Bruns, T., Lee, S., & Taylor, J. (1990). Amplification and direct sequencing of fungal

ribosomal RNA genes for phylogenetics. *PCR Protocols: A Guide to Methods and Applications*, *3*, 315–322.

Wu, P. K., Tai, W. C. S., Choi, R. C. Y., Tsim, K. W. K., Zhou, H., Liu, X., … Hsiao, W. L. W. (2011). Chemical and DNA authentication of taste variants of *Gynostemma pentaphyllum* herbal tea. *Food Chemistry*, *128*, 70–80.

Zhang, Z., Schwartz, S., Wagner, L., & Miller, W. (2000). A greedy algorithm for aligning DNA sequences. *Journal of Computational Biology*, *7*, 203–214.

Zheng, S., Jiang, X., Wu, L., Wang, Z., & Huang, L. (2014). Chemical and Genetic Discrimination of Cistanches Herba Based on UPLC-QTOF / MS and DNA Barcoding. *PLoS ONE*, *9*. https://doi.org/10.1371/journal.pone.0098061

Figure captions

Fig. 1. Representative profiles of the four *Carduus* species (λ=330nm). Compounds are numbered as in Table 1. (A) *Carduus argyroa*; (B) *Carduus cephalanthus*; (C) *Carduus pycnocephalus*; (D) *Carduus nutans* subsp*. macrocephalus*

Fig. 2. Statistical analysis of the 39 samples from the four *Carduus* species based on the 31 *Carduus* target compounds as variables. (A, B, C, D) PCA score and loading plots of samples from the four *Carduus* species. (E) Hierarchical cluster analysis.

Fig. 3. Comparison of ITS sequences among the four *Carduus* species. (A) Structure of the rRNA gene cluster, arrows indicates the position of the primers (ITS1 and ITS4). (B, C) *Carduus* ITS sequence alignment is shown. Variations in the nucleotide composition among the species are shaded. *Nde*I, *Stu*I and *Apa*LI sites are indicated by squared solid box, dotted box, and dashed box, respectively.

Fig. 4. Hierarchical cluster analysis of *Carduus* species (numbers at the node indicate bootstrap values). (A) Neighbor Joining tree. (B) UPGMA tree.

Fig. 5. Capillary gel electrophoresis analysis of PCR products of the four *Carduus* species ITS regions and the relative fragments produced by the *Nde*I, *Stu*I and *Apa*LI restriction enzymes. Lanes: 1, a single product of about 811 bp is produced by *C. argyroa*; 2 a single product of about 807 bp is produced by *C. nutans* subsp*. macrocephalus;* 3, a single product of about 798 bp is produced by *C. pycnocephalus*; 4, a single product of about 785 bp is produced by *C. cephalanthus*; 5, *C. argyroa* PCR products are not digested by *Nde*I; 6, *C. nutans* subsp*. macrocephalus* PCR products digested

by *Nde*I give two major fragments of 512 and 295 bp, respectively; 7, *C. pycnocephalus* PCR products are not digested by *Nde*I; 8, *C. cephalanthus* PCR products are not digested by *Nde*I; 9, *Stu*I cleaves *C. argyroa* ITS region, giving two fragments 623 and 188 bp long; 10, *C. nutans* subsp. *macrocephalus* PCR products are not digested by *Stu*I; 11, *Stu*I cleaves *C. pycnocephalus* ITS region, giving two fragments, 620 and 178 bp long; 12, *C. cephalanthus* PCR products are not digested by *Stu*I; 13, *C. argyroa* PCR products are not digested by *Apa*LI; 14, *C. nutans* subsp. *macrocephalus* PCR products are not digested by *Apa*LI; 15, PCR products from *C. pycnocephalus* give two fragments, of 508 and 290 bp, when digested by *Apa*LI; 16, *C. cephalanthus* PCR products are not digested by *Apa*LI.

Supplementary material

Table S1. Sites and date of collection, voucher specimens, and GenBank accession numbers of the four *Carduus* species

Table S2. Multiple Reaction Monitoring acquisitions

Table S3. Sequence homologies of ITS sequences between each pair of *Carduus* sequences

Fig. S1. Three dimensional PCA score plot of samples of the four *Carduus* species.

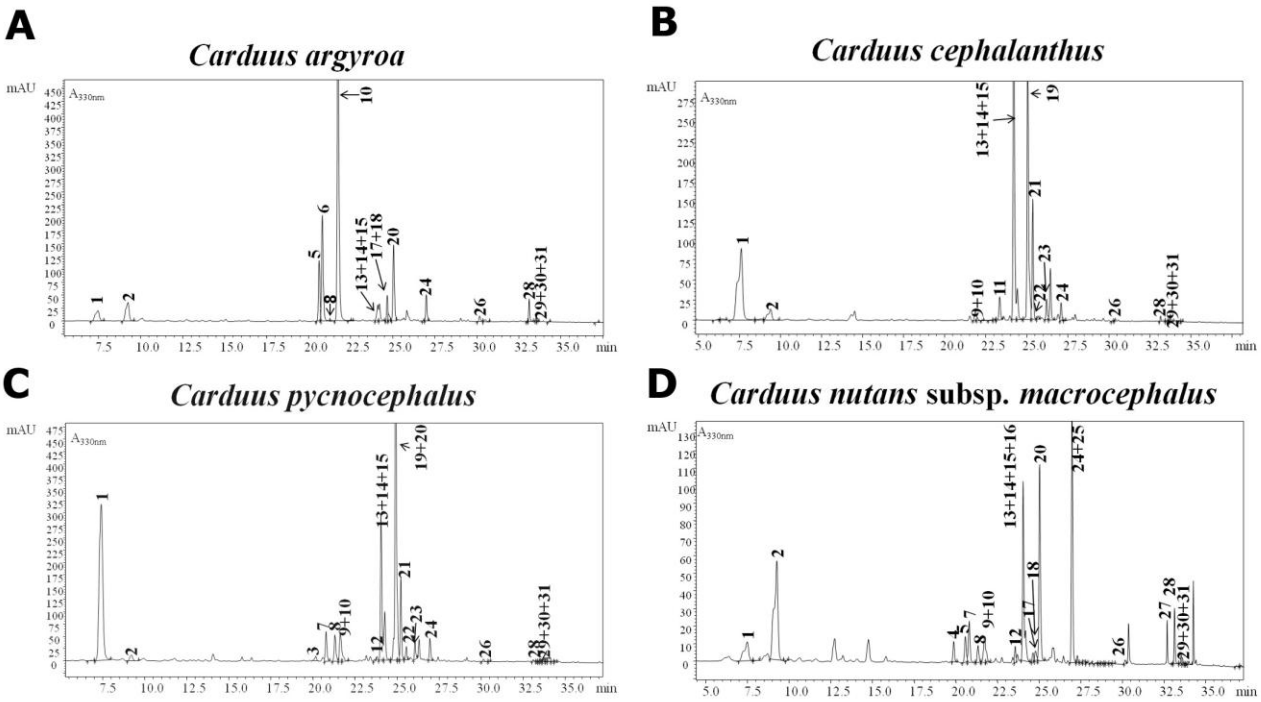
A
Carduus argyroa
mAU
A330nm
1
2
5
6
8
10
13+14+15
17+18
20
24
26
28
29+30+31
min
B
Carduus cephalanthus
mAU
A330nm
1
2
9+10
11
13+14+15
19
21
22
23
24
26
28
29+30+31
min
C
Carduus pycnocephalus
mAU
A330nm
1
2
3
7
8
9+10
12
13+14+15
19+20
21
22
23
24
26
28
29+30+31
min
D
Carduus nutans subsp. macrocephalus
mAU
A330nm
1
2
4
5
7
8
9+10
12
13+14+15+16
17
18
20
24+25
26
27
28
29+30+31
min

Fig.1

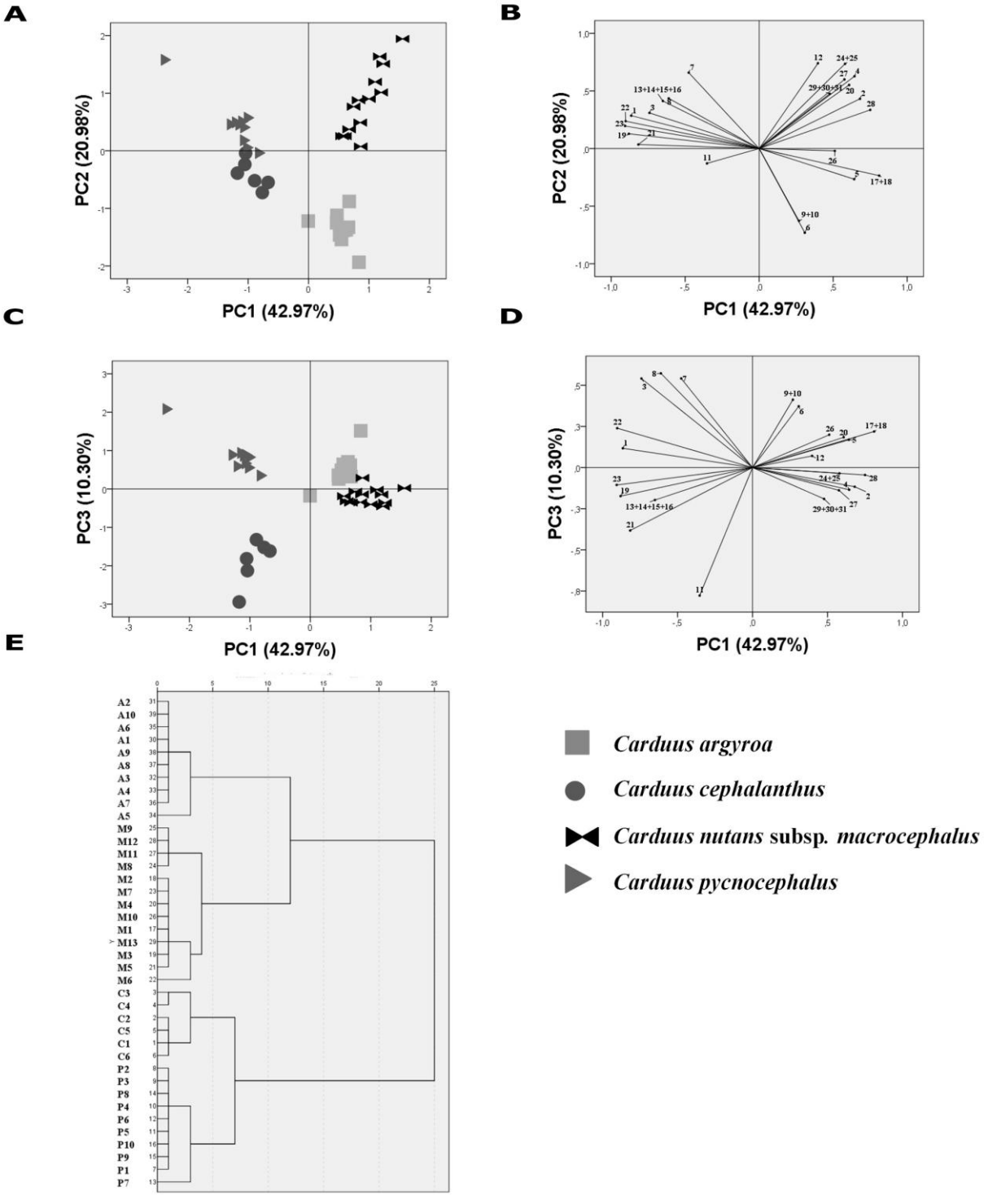
A
PC2 (20.98%)
PC1 (42.97%)
B
PC2 (20.98%)
PC1 (42.97%)
C
PC3 (10.30%)
PC1 (42.97%)
D
PC3 (10.30%)
PC1 (42.97%)
E
A2
A10
A6
A1
A9
A8
A3
A4
A7
A5
M9
M12
M11
M8
M2
M7
M4
M10
M1
M13
M3
M5
M6
C3
C4
C2
C5
C1
C6
P2
P3
P8
P4
P6
P5
P10
P9
P1
P7
Carduus argyroa
Carduus cephalanthus
Carduus nutans subsp. macrocephalus
Carduus pycnocephalus

Fig.2

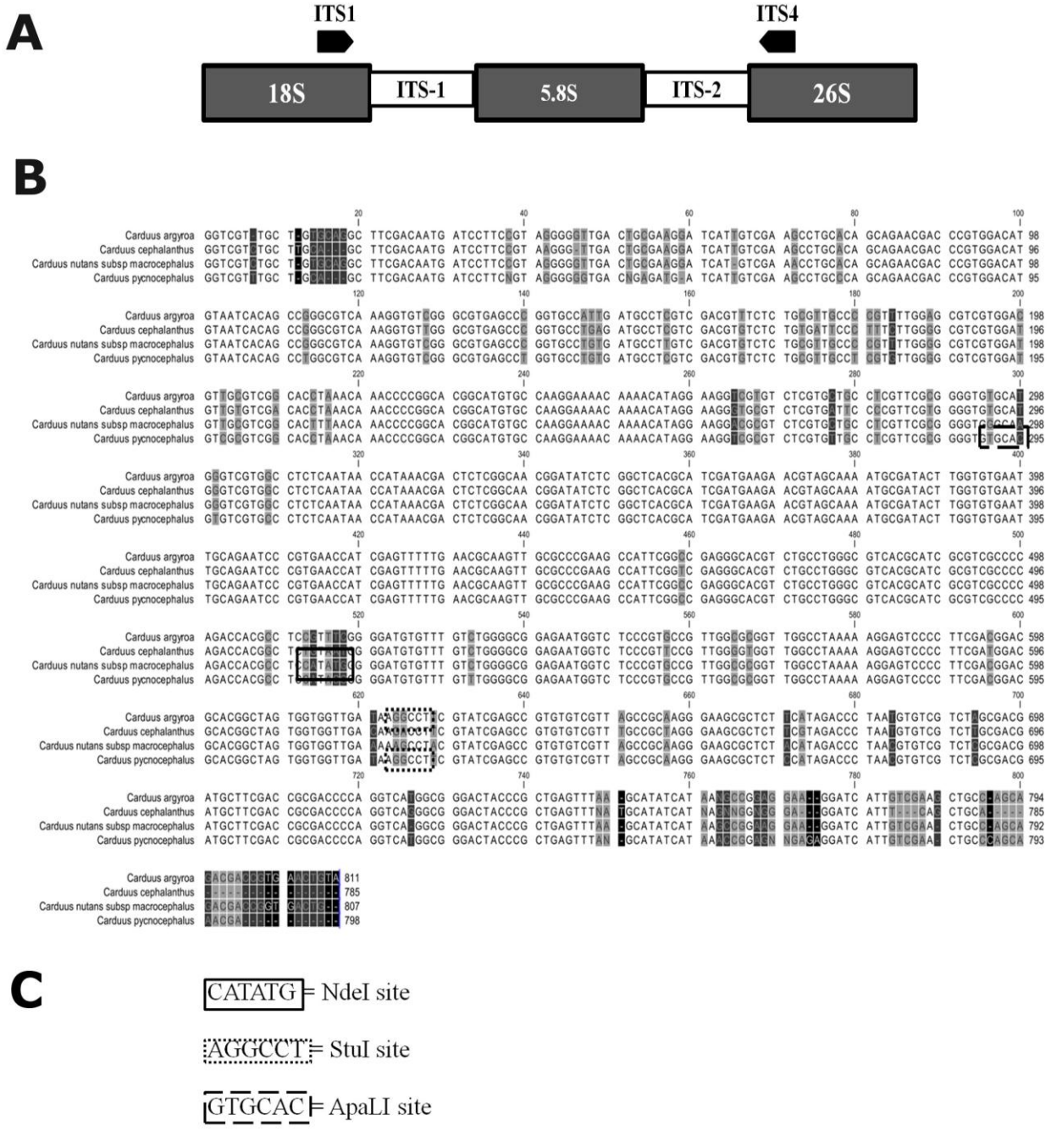
A
ITS1
ITS4
18S
ITS-1
5.8S
ITS-2
26S
B
Carduus argyroa
Carduus cephalanthus
Carduus nutans subsp macrocephalus
Carduus pycnocephalus
C
CATATG = NdeI site
AGGCCT = StuI site
GTGCAC = ApaLI site

Fig.3

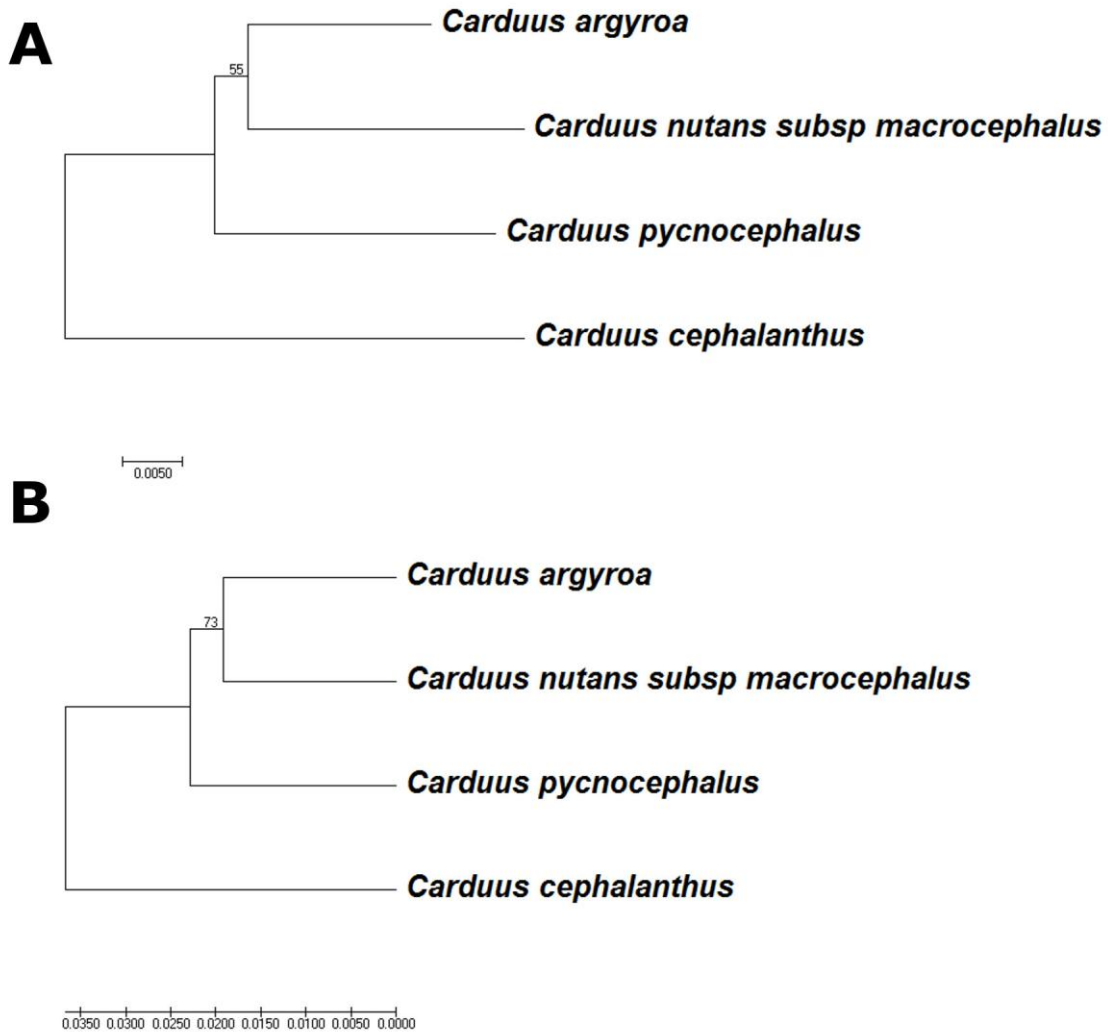
A
Carduus argyroa
55
Carduus nutans subsp macrocephalus
Carduus pycnocephalus
Carduus cephalanthus
0.0050
B
Carduus argyroa
73
Carduus nutans subsp macrocephalus
Carduus pycnocephalus
Carduus cephalanthus
0.0350 0.0300 0.0250 0.0200 0.0150 0.0100 0.0050 0.0000

Fig.4

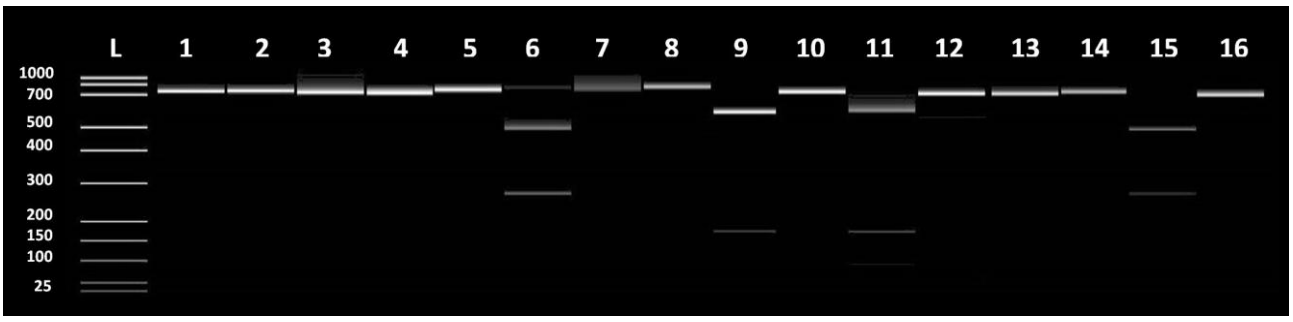
L
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
1000
700
500
400
300
200
150
100
25

Fig.5

Table 1. List of identified and putatively-identified compounds in the four *Carduus* species. Each compound is quoted through its relative retention time, UV spectrum, molecular formula, pseudomolecular ions, molecular weight fragments obtained by Product Ion Scan mode (PIS), identified or tentatively-identified compound names. The identification confidence value, the presence of the compounds in the *Carduus* species, and the references are also given.

N°	RT (min)	λ_{max} (nm)	Putative molecular formula	$[M-H]^+$ m/z	$[M-H]^-$ m/z	Mol. weight g/mol	MS^2+ m/z	MS^2- m/z	Compound Name	IC [b]	Presence in *Carduus* species[C]	References
1[a]	7.282	325	$C_{16}H_{18}O_9$	355	353	354	163	191	5-*O*-Caffeoylquinic acid (chlorogenic acid)	1	A, C, M, P	(Dimitrova-dyulgerova et al., 2015; Li et al., 2014; Marrelli et al., 2013)
2[a]	9.075	325	$C_{16}H_{18}O_9$	355	353	354	163	191	4-*O*-Caffeoylquinic acid (Cryptochlorogenic acid)	1	A, C, M, P	(Li et al., 2014)
3	19.88	255 353	$C_{27}H_{30}O_{16}$	611	609	610	303, 121, 127, 135, 153, 161	301	Quercetin diglycoside	2	P	(Dimitrova-dyulgerova et al., 2015; Jordon-Thaden & Louda, 2003; Li et al., 2014)
4	19.89	348 254 265	$C_{27}H_{30}O_{16}$	611	609	610	287 135 137 153 161	285	Luteolin *O*-diglucoside	2	M	(Bain & Desrochers, 1988; Dimitrova-dyulgerova et al., 2015; Jordon-Thaden & Louda, 2003)
5	20.617	348 253 265	$C_{26}H_{30}O_{16}$	581	579	580	287 135 137 153	285	Luteolin *O*-arabinosyl-glucoside	2	A, M	(Cuyckens & Claeys, 2004; Li et al., 2014)
6	20.815	348 253 265	$C_{26}H_{30}O_{16}$	581	579	580	287 135 137 153	285	Luteolin *O*-arabinosyl-glucoside isomer	2	A	(Cuyckens & Claeys, 2004; Li et al., 2014)
7[a]	20.823	255 351	$C_{27}H_{30}O_{16}$	611	609	610	303	301	Quercetin 3-*O*-rutinoside (rutin)	1	M, P	(Dimitrova-dyulgerova et al., 2015; Li et al., 2014; Marrelli et al., 2013)
8[a]	21.351	255 351	$C_{21}H_{20}O_{12}$	465	463	464	303	301	Quercetin 3-*O*-glucoside	1	A, M, P	(Li et al., 2014)

9[a]	21.651	348 253 265	$C_{21}H_{18}O_{12}$	463	461	462	287 135 137 153 161	285	Luteolin 7-*O*-glucuronide	1	C, M, P	(Cuyckens & Claeys, 2004)
10[a]	21.758	349 253 265	$C_{21}H_{20}O_{11}$	449	447	448	287 135 137 153 161	285	Luteolin 7-*O*-glucoside	1	A, C, M, P	(Cuyckens & Claeys, 2004; Li et al., 2014)
11	23.233	329	$C_{25}H_{24}O_{12}$	517	515	516	163	191	Dicaffeoylquinic acid	3	C	(Li et al., 2014)
12[a]	23.702	265 344	$C_{27}H_{30}O_{15}$	595	593	594	287	285	Kaempferol 3-*O*-rutinoside	1	M, P	(Li et al., 2014)
13[a]	24.109	329	$C_{25}H_{24}O_{12}$	517	515	516	163	191	1,5-dicaffeoylquinic acid	1	A, C, M, P	(Li et al., 2014)
14[a]	24.231	265 344	$C_{21}H_{20}O_{11}$	449	447	448	287	285	Kaempferol-3-*O*-glucoside	1	A, C, M, P	(Li et al., 2014)
15[a]	24.257	329	$C_{25}H_{24}O_{12}$	517	515	516	163	191	3,5-dicaffeoylquinic acid	1	A, C, M, P	(Li et al., 2014)
16	24.257	348 253 266		653	651	652	449 329 287 135 137 153 161	609 447 327 285	Luteolin acetyl diglycoside	3	M	(Cuyckens & Claeys, 2004)
17	24.7	337 266	$C_{27}H_{30}O_{16}$	593	591	592	271 119 153 163	269	Apigenin *O*-rhamnosyl-glucuronide	2	A, M	(Jordon-Thaden & Louda, 2003)
18[a]	24.836	337 266	$C_{21}H_{20}O_{10}$	433	431	432	271	269	Apigenin 7-*O*-glucoside	1	A, M	(Li et al., 2014)
19	24.916	344 252 266	$C_{27}H_{34}O_{17}$	595	593	594	301	299	Diosmetin arabinosyl-glucoside	2	C, E	(Li et al., 2014)

20[a]	25.074	337 266	$C_{21}H_{20}O_{10}$	447	445	446	271	269	Apigenin 7-*O*-glucuronide	1	A, M, P	(Li et al., 2014)
21	25.221	344 251 266	$C_{27}H_{34}O_{17}$	595	593	594	301	299	Diosmetin arabinosyl-glucoside	2	C, P	(Li et al., 2014)
22[a]	25.552	338 267	$C_{28}H_{32}O_{15}$	609	607	608	301	299	Diosmetin *O*-rhamnosyl-glucoside (diosmin)	1	C, P	(Li et al., 2014)
23[a]	26.061	329	$C_{25}H_{24}O_{12}$	517	515	516	163	191	4,5-dicaffeoylquinic acid	1	C, P	(Li et al., 2014)
24[a]	27.037	263 343	$C_{21}H_{20}O_{10}$	433	431	432	287	285	Kaempferol 3-*O*-rhamnoside	1	A, C, M, P	(Al-Shammari et al., 2015; Jordon-Thaden & Louda, 2003)
25	27.037	263 343		667	665	666	463 343 301 286	623 461 341 299 284	Diosmetin acetyl diglycoside	3	M	(Li et al., 2014)
26[a]	30.232	349 252 266	$C_{15}H_{10}O_6$	287	285	286	135 137 153 161		Luteolin	1	A, C, M, P	(Bain & Desrochers, 1988; Dimitrova-dyulgerova et al., 2015; Jordon-Thaden & Louda, 2003; Li et al., 2014)
27	32.678	347 266		709	707	708	343 301 286	665 341 299 284	Diosmetin diacetyl diglycoside	3	M	(Li et al., 2014)
28[a]	33.168	337 266	$C_{15}H_{10}O_5$	271	269	270	119 153 163		Apigenin	1	A, C, M, P	(Al-Shammari et al., 2015; Dimitrova-dyulgerova et al., 2015; Li et al., 2014)
29[a]	33.569	366 265	$C_{15}H_{10}O_6$	287	285	286	121 137 153 165		Kaempferol	1	A, C, M, P	(Al-Shammari et al., 2015; Dimitrova-dyulgerova et al., 2015; Jordon-Thaden & Louda, 2003; Li et al., 2014)
30[a]	33.64	345 267	$C_{16}H_{12}O_6$	301	299	300	258 286	256 284	Diosmetin	1	A, C, M, P	(Li et al., 2014)
31[a]	33.64	266	$C_{17}H_{14}O_7$	331	329	330	153	161	Tricin	1	A, C, M, P	(Li et al., 2014)

		348					203 315					

[a] Compounds identified by comparison with reference standards.

[b] The Identification Confidence value is in agreement with the CAWG (2007) guidelines and indicates: Level 1: Identified compound (A minimum of two independent orthogonal data (such as retention time and mass spectrum) compared directly to an authentic reference standard; Level 2: Putatively annotated compound (similarity of chromatographic and spectral data to published data); Level 3: Putatively characterized class of compounds

[C] A= *C. argyroa*; C= *C. cephalanthus*; M= *C. nutans* subsp. *macrocephalus*; P= *C. pycnocephalus*.

Supplementary material

Table S1 Sites and date of collection, voucher specimens, and GenBank accession numbers of the four *Carduus* species

Species	**Sites and date of collection**	**Coordinates**	**Voucher speciment**	**GenBank Access. No.**
Carduus argyroa	Decimomannu, 27 May 2015	(39°17'47.96"N - 8°58'14.95"E)	CAG-803	KY242483
Carduus cephalanthus	Capo Testa, 12 June 2015	(41°14'33.80"N – 9°8'49.25"E)	CAG-807	KY242486
Carduus nutans subsp. *macrocephalus*	Gennargentu, 18 June 2015	(39°57'35.77"N - 9°19'12.46"E)	CAG-802	KY242485
Carduus pycnocephalus	Monte dei Sette Fratelli, 21 May 2015	(39°20'43.60"N – 9°17'43.74"E)	CAG-805	KY242484

Table S2. Multiple Reaction Monitoring acquisitions

Compound	**MRM (m/z) ESI$^+$**	**MRM (m/z) ESI$^-$**
Chlorogenic acid	355→163	353→191
Dicaffeoylquinic acid	517→163	515→191
Apigenin	271→119 271→153 271→163	
Luteolin	287→127 287→135 287→137 287→153 287→161	
Kaempferol	287→121 287→137 287→153 287→165	
Quercetin	303→121 303→127 303→137 303→153 303→161 303→165	
Diosmetin	301→258 301→286	299→256 299→284

Table S3. Sequence homologies of ITS sequences between each pair of *Carduus* sequences are shown

	C. argyroa	***C. cephalanthus***	***C. nutans* subsp. *macrocephalus***	***C. pycnocephalus***
C. argyroa	100%	93%	95%	94%
C. cephalanthus	93%	100%	92%	92%
C. nutans* subsp. *macrocephalus	95%	92%	100%	94%
C. pycnocephalus	94%	92%	94%	100%

Fig. S1. Three dimensional PCA plot of samples of the four *Carduus* species.